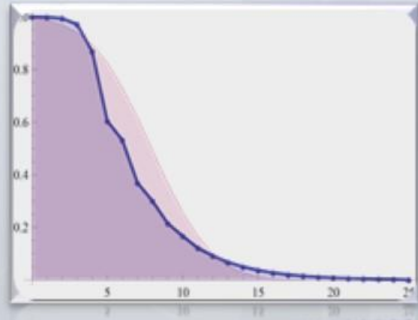
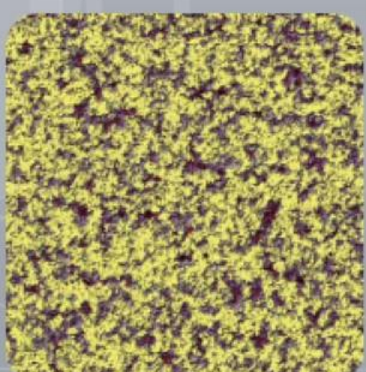


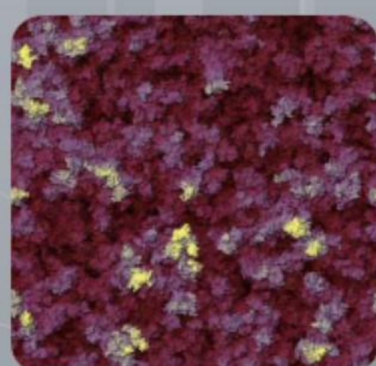
Hebrew University

Math Colloquium

December 2012

Cutoff Phenomenon: Instant Randomness



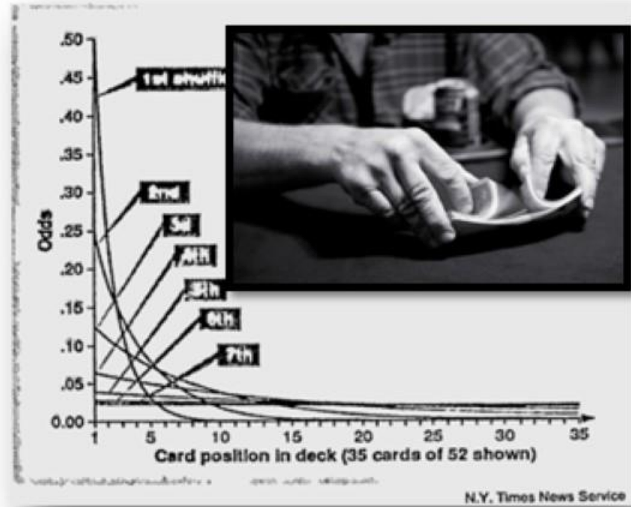


Eyal Lubetzky

Microsoft Research

Shuffling cards

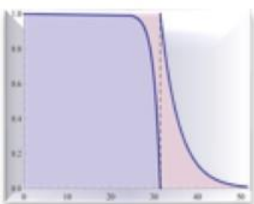
- How many shuffles are needed to mix a deck of cards? (e.g., can we say where A♥ is, does it precede K♣, …)
  - [Aldous, Diaconis '86]:
    "For card players, the question is not 'exactly how close to uniform is the deck after a million riffle shuffles?', but 'is 7 shuffles enough?'"
  - *Is there a sharp transition (**cutoff**)?*

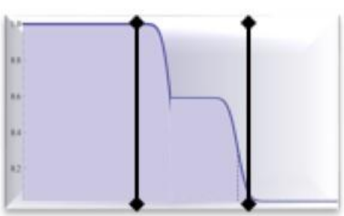


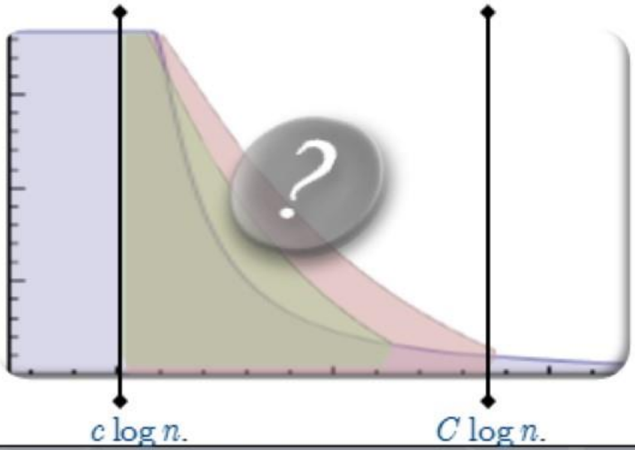
- Formally: understand the mixing time (t_{mix}) of the random walk on the symmetric group with a prescribed set of generators (e.g., all transpositions).

Walking on groups

- What is t_{mix} of the RW on the Cayley graph $(\Gamma, S \cup S^{-1})$ for $\Gamma = \mathrm{PSL}_2(\mathbb{F}_q)$ and $S = \left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \right\}$?
  - d-regular *expander* for $d = 4$ (the spectrum of the adjacency matrix supported on $(-d+\varepsilon, d-\varepsilon) \cup \{d\}$).
- On any expander: rapid convergence to equilibrium: within $[c \log n, C \log n]$ (not "too gradual").
  - *Is there cutoff (convergence described by step function)?*
  - *Multiple step functions?*





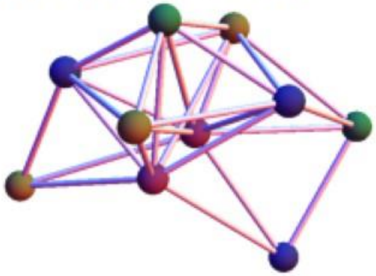


Sampling a coloring

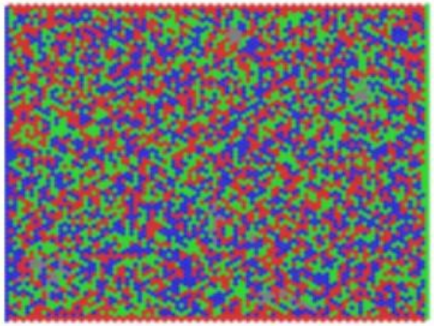
- Goal: given a graph $G = (V, E)$ and $q \geq 2$, sample a legal q-coloring uniformly.

N. Metropolis
(1915 – 1999)

- Metropolis algorithm:
  - Select random vertex and random color.
  - Accept the change if legal.
- *Is there cutoff?*



- Relaxation: q-state Potts model with $\lambda \in [0, \infty]$.
  - Select random vertex.
  - Change it to color $x \in [q]$ with probability $\propto \lambda^{\#\{\text{neighbors with color } x\}}$



Measuring convergence

- Standard choice: L^1 (total-variation) mixing time to within ε of the stationary distribution π :

$$t_{\text{mix}}(\varepsilon) = \inf\left\{t : \max_x \left\|P^t(x,\cdot) - \pi\right\|_{\text{TV}} \le \varepsilon\right\}.$$

where

$$\left\|\mu - \nu\right\|_{\text{TV}} = \sup_{A\subset\Omega}\left[\mu(A) - \nu(A)\right].$$

- Monotone decreasing and decays exponentially:

$$\max_x \left\|P^t(x,\cdot) - \pi\right\|_{\text{TV}} \le 2^{-\ell} \quad \text{for} \quad t \ge \ell\, t_{\text{mix}}\left(\tfrac{1}{4}\right)$$

The Cutoff Phenomenon

- Describes a sharp transition in the convergence of finite ergodic Markov chains to stationarity.

Steady convergence
it takes a while to reach distance ½ from stationarity then a while longer to reach distance ¼, etc.

Abrupt convergence
distance from equilibrium quickly drops from 1 to 0

Cutoff: formal definition

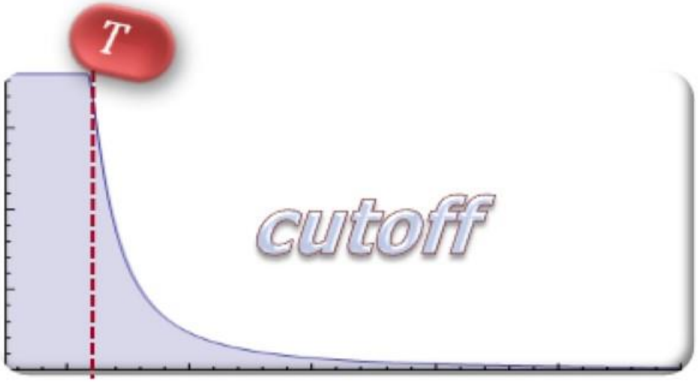
- Standard notion of convergence: L^1 (total variation) .
  - $d_{\mathrm{TV}}(t) = \max_{x\in\Omega} \sup_{A\subset\Omega} \left|\mathbb{P}_x(X_t \in A) - \pi(A)\right|$
  - $t_{\mathrm{mix}}(\varepsilon) = \min\left\{ t : d_{\mathrm{TV}}(t) < \varepsilon \right\}$
  - Convention: define *mixing time* as $t_{\mathrm{mix}}(1/e)$.
- A family of chains $\left(X_t^n\right)$ is said to have ***cutoff*** if:

$$\lim_{n\to\infty} \frac{t_{\mathrm{mix}}(\varepsilon)}{t_{\mathrm{mix}}(1-\varepsilon)} = 1 \quad \forall\ 0 < \varepsilon < 1.$$

i.e., $t_{\mathrm{mix}}(\alpha) = (1+o(1))\,t_{\mathrm{mix}}(\beta)$ for any $0 < \alpha , \beta < 1$.

Cutoff: widespread phenomenon?

- Take an MCMC sampler that mixes $\asymp T(n)$ steps.
- Cutoff at time T :
  - At time $(1-\varepsilon)T$ the distance to stationarity is $1 - o(1)$.
  - At time $(1+\varepsilon)T$ that distance becomes $o(1)$.



- Many natural chains are *conjectured* to have cutoff, e.g., spins-systems at high temperatures:

{ **Ising** on lattices ; **Potts** model on lattices;
Gas Hard–core model on lattices; lattice **Colorings** ;
Anti–ferromagnetic Ising / Potts model, **Spin–glass**,
Arbitrary boundary conditions / external field; ... }

yet proving cutoff can be extremely challenging.

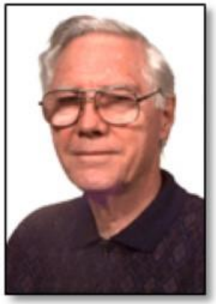
Cutoff History

D. Aldous

P. Diaconis

- Discovered:
  - Random transpositions on S_n [Diaconis, Shahshahani '81]
  - RW on the $\mathbb{Z}_2^n$, riffle-shuffle [Aldous '83]
- Notable examples:
  - Top-in-at-random shuffle [Aldous, Diaconis '86]
  - Riffle-shuffle [Bayer, Diaconis '92]
  - RWs on finite groups [Saloff-Coste '04]
  - 1D Markov chains: Birth-and-Death chains [Diaconis, Saloff-Coste '06], [Ding, L., Peres '09]
- Nearly 3 decades after its discovery:
  - *Unknown*: cutoff for RW on an expander? *uniform π*
  - *Unknown*: cutoff for spin-systems on lattices? *intractable π*

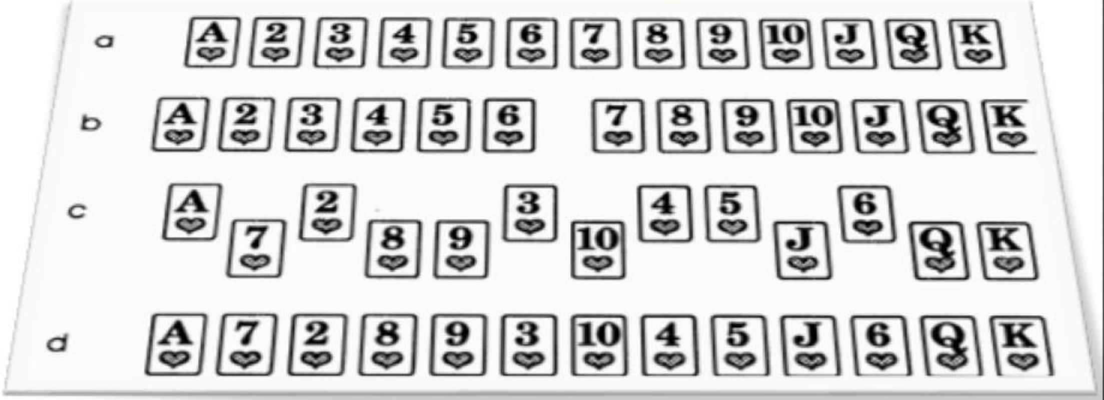
The riffle-shuffle



E. Gilbert

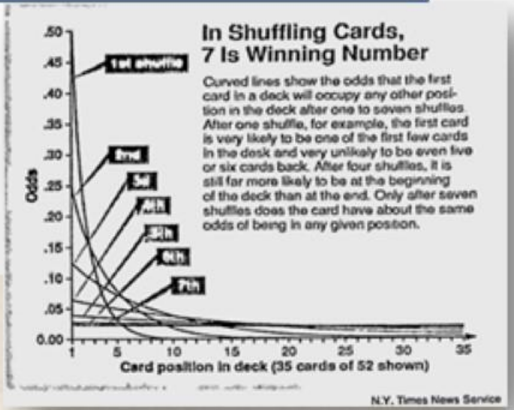
C.E. Shannon (1916 – 2001)

- Modeled by Gilbert-Shannon (1955) and independently Reeds (1981):
  - Cut $m \sim \mathrm{Bin}(n, \frac{1}{2})$ cards from the top.
  - Riffle the two packs together:
    - With l, r cards in *left* and *right* hands, resp., drop card from *left* with probability $\frac{l}{l+r}$.



(Illustration from [Bayer-Diaconis '92])

‘7 shuffles suffice’



The New York Times January 9, 1990

It takes just seven ordinary, imperfect shuffles to mix a deck of cards thoroughly, researchers have found. Fewer are not enough and more do not significantly improve the mixing.

The mathematical proof, discovered after studies of results from elaborate computer calculations and careful observation of card games, confirms the intuition of many gamblers, bridge enthusiasts and casual players that most shuffling is inadequate.

...

By saying that the deck is completely mixed after seven shuffles, Dr. Diaconis and Dr. Bayer mean that every arrangement of the 52 cards is equally likely or that any card is as likely to be in one place as in another.

The cards do get more and more randomly mixed if a person keeps on shuffling more than seven times, but seven shuffles is a transition point, the first time that randomness is close. Additional shuffles do not appreciably alter things...

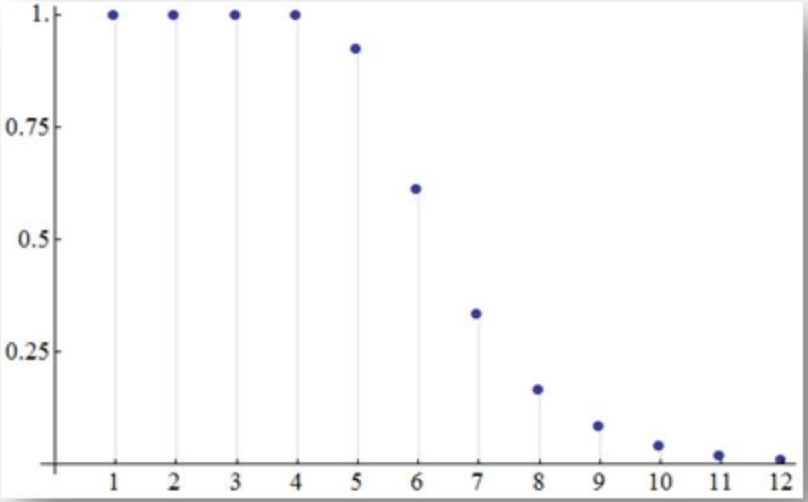
'7 shuffles suffice' (ctd.)

- [Bayer, Diaconis '92]: numerically computed the distance to stationarity after k riffle-shuffles:

1	2	3	4
1.	1.	1.	1.

5	6	7	8
0.9237	0.6135	0.3341	0.1672

9	10	11	12
0.0854	0.0429	0.0215	0.0108



Useful tool: strong stationary times

- DEFINITION: A *strong stationary time* for a Markov chain (X_t) with stationary measure π is a randomized stopping time τ such that $X_\tau \sim \pi$ independent of τ, *i.e.*

$$\forall t : \mathbb{P}(\tau = t,\, X_\tau = y) = \mathbb{P}(\tau = t)\,\pi(y).$$

$$(\;\Leftrightarrow\; \forall t : \mathbb{P}(\tau \le t,\, X_t = y) = \mathbb{P}(\tau \le t)\,\pi(y).\;)$$

- THEOREM: ([Aldous-Diaconis '86,'87]

Let τ be a strong stationary time for a Markov chain (X_t) with stationary distribution π and let t_0 be an integer such that $\max_{x\in\Omega} \mathbb{P}_x(\tau > t_0) \le \varepsilon$. Then $t_{\mathrm{mix}}(\varepsilon) \le t_0$.

Useful tool: strong stationary times (ctd.)

- RECALL: τ is a *strong stationary time* for (X_t) w.r.t. π if
$$\mathbb{P}(\tau \le t,\ X_t = y) = \mathbb{P}(\tau \le t)\,\pi(y). \qquad (\ast)$$
- Upper bound on mixing:

τ strong stationary time for (X_t) ,
$$\max_{x\in\Omega} \mathbb{P}_x(\tau > t_0) \le \varepsilon \implies t_{\mathrm{mix}}(\varepsilon) \le t_0$$

- PROOF:

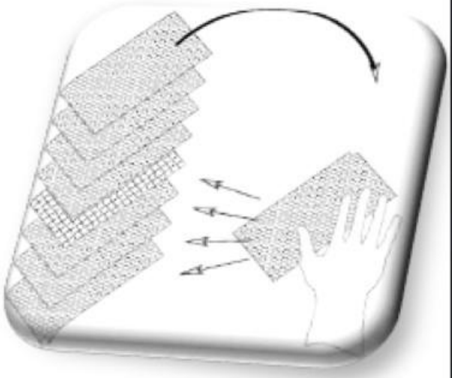
Since
$$\mathbb{P}_x(X_{t_0} \in A) \le \mathbb{P}_x(\tau \le t_0\ ,\ X_{t_0} \in A) + \mathbb{P}_x(\tau > t_0),$$
plugging in $(\ast)$ gives
$$\sup_A \left[\mathbb{P}_x(X_{t_0} \in A) - \pi(A)\right] \le \mathbb{P}_x(\tau > t_0). \qquad \blacksquare$$

Top-to-random shuffle

- Strong stationary time: 1 step after bottom reaches top:

$$\tau = \min\left\{t : \sigma_t(1) = n\right\} + 1.$$

  - Proof: internal ordering of the cards below card #n is uniform (induction).



- Similarly to the coupon collector: $\tau = \tau_1 + \tau_2 + \ldots + \tau_{n-1} + 1$ for $\tau_i \sim \mathrm{Geom}(k/n)$ ind.
- <u>COROLLARY</u>:

$$t_{\mathrm{mix}}(\varepsilon) \le n \log n + \log(\tfrac{1}{\varepsilon})n\,.$$

- Tight: as long as the *j*-th card from bottom did not reach the top, last *j* cards retain internal ordering…

Cutoff for shuffles

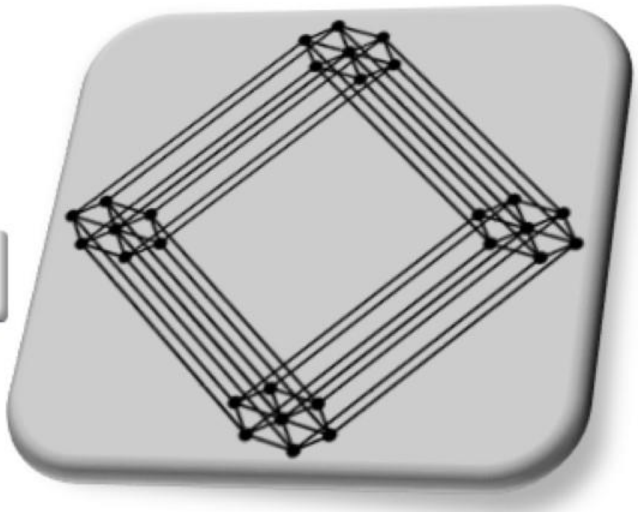
- Sharp mixing results for the shuffles discussed so far:

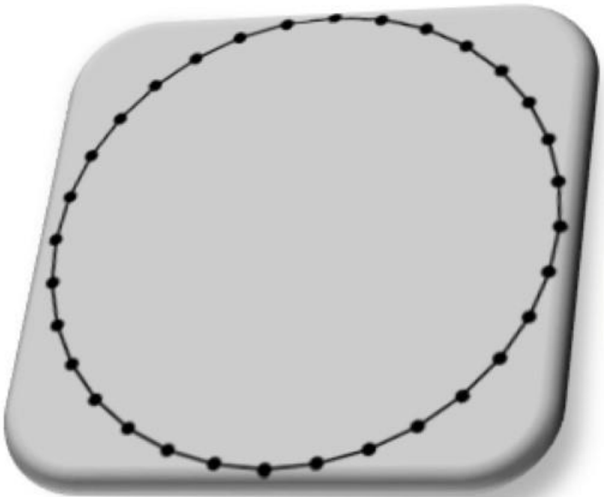
Shuffle	Cutoff	Ref.
Random transpositions	$\frac{1}{2} n \log n$	[Diaconis, Shahshahani ’81]
Riffle shuffle	$\frac{3}{2} n \log n$	[Aldous ’83], [Bayer, Diaconis ‘92]
Top–to–random	$n \log n$	[Aldous, Diaconis ‘86]

- What about random-to-random?
  ? *Unknown whether or not there is cutoff…*

Basic examples: RWs on graphs

Lazy discrete-time simple random walk





On the hypercube $\{0,1\}^n$:
☑ Exhibits cutoff at
$c_0\, n \log n + O(n)$
[Aldous '83]

On the n–cycle:
☒ No cutoff.

RW on the hypercube

- Let (X_t) be a lazy simple RW on the hypercube $\{0,1\}^n$.
- Each step: select coordinate $J_t \in [n]$ and update $I_t \in \{0,1\}$ both independent uniform.
- Strong stationary time:

$$\tau_{\text{refresh}} = \min\left\{t : \{J_1, \ldots, J_t\} = [n]\right\}.$$

- By the coupon collector paradigm:

$$\max_{x \in \Omega} \mathbb{P}_x\left(\tau_{\text{refresh}} > n \log n + cn\right) \le \mathrm{e}^{-c},$$

thus

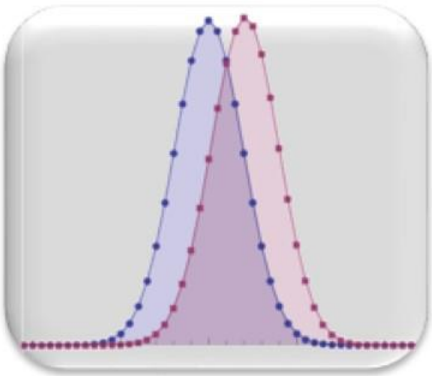
$$t_{\text{mix}}(\varepsilon) \le n \log n + \log(\tfrac{1}{\varepsilon}) n.$$

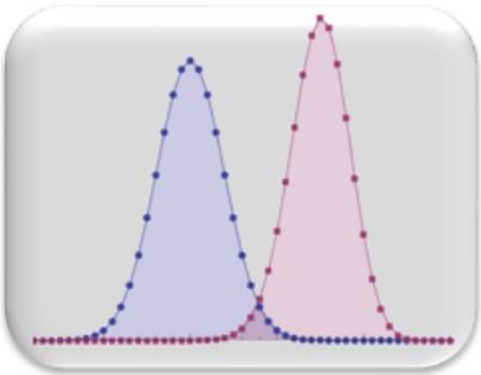
- From below: general lower bound of $\frac{1}{2} n \log n$ applies…

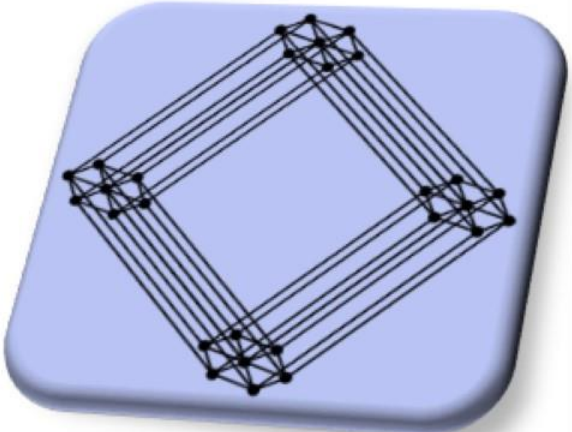
RW on the hypercube

Choose coordinate $\{1,\dots,n\}$ and new $\{0,1\}$ value for it, uniformly & independently.

- [Aldous '83]: cutoff at $\frac{1}{2}n\log n$ with window $O(n)$:
  - Symmetry: start at the all-1 state.
  - # of 1's at time t is ~ $\mathrm{Bin}\left(n,(1+e^{-t/n})/2\right)$.
  - # of 1's under stationary measure ~ $\mathrm{Bin}(n,½)$, which has Gaussian fluctuations of $O(\sqrt{n})$.
  - Mixing occurs when $e^{-t/n} \asymp \sqrt{n}$ (match fluctuations).

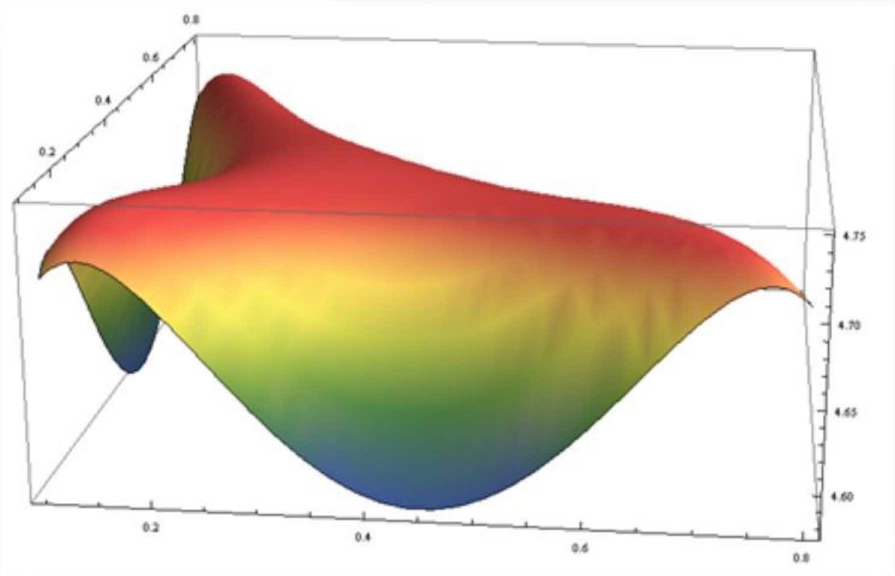






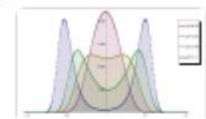
From the hypercube to Potts

- RECALL:
  - Metropolis for the q-state Potts model with $\lambda \in [0,\infty]$:
    select a random vertex and assign it the color $j \in [q]$
    with probability $\propto \lambda^{\#\{\text{neighbors with color } j\}}$.
  - Lazy RW on $[q]^n$: $\equiv$ *Potts model with* $\lambda = 1$



Cutoff for spin–systems

- Till recently: *only* spin-systems where cutoff was verified are Ising and Potts models on the *complete graph* [Levin, Luczak, Peres '10], [Ding, L., Peres '09], [Cuff, Ding, L., Louidor, Peres, Sly '12]



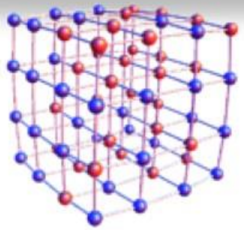
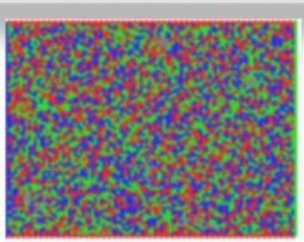
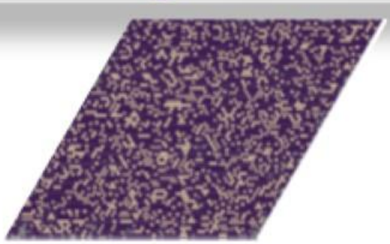
- Conjectured to believe at high temperatures for:
  - Ising on the lattice.
  - Potts model on the lattice.
  - Independent sets in lattices.
  - Colorings of lattices.

Unknown even in 1 dimension (Q. of Peres)...

- Recently: [L., Sly '13+] settled Ising on the torus $\mathbb{Z}_n^d$.
- New results [L., Sly '14+] bypass symmetry/monotonicity constraints and proves cutoff in general spin-systems.

New cutoff results (ctd.)

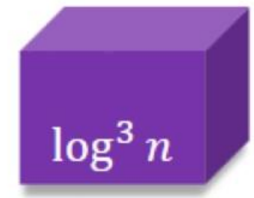
- Examples for some of the models on the **lattice** $\mathbb{Z}_n^d$ with arbitrary boundary conditions:
  - Proper coloring with $q \geq 4d(d+1)$ colors
  - Potts model: $q \geq 2$ colors and $1 \leq \lambda < \left(1+\frac{q}{2d}\right)^{-4d}$
  - Independent sets on lattices, Ising on arbitrary geometries with sub-exponential growth, Anti-ferromagnetic Potts,…

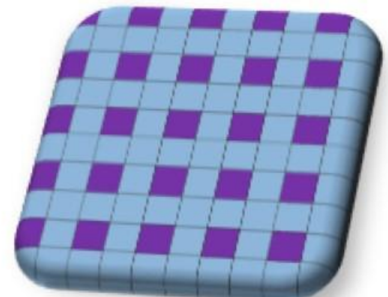
- ? Q: cutoff for Metropolis for legal colorings on a 3-regular transitive **expander**? On a random 3-regular graph?

Intuition: cutoff on the lattice

- Break up $\mathbb{Z}_n^d$ to cubes of side-length $\log^3 n$.

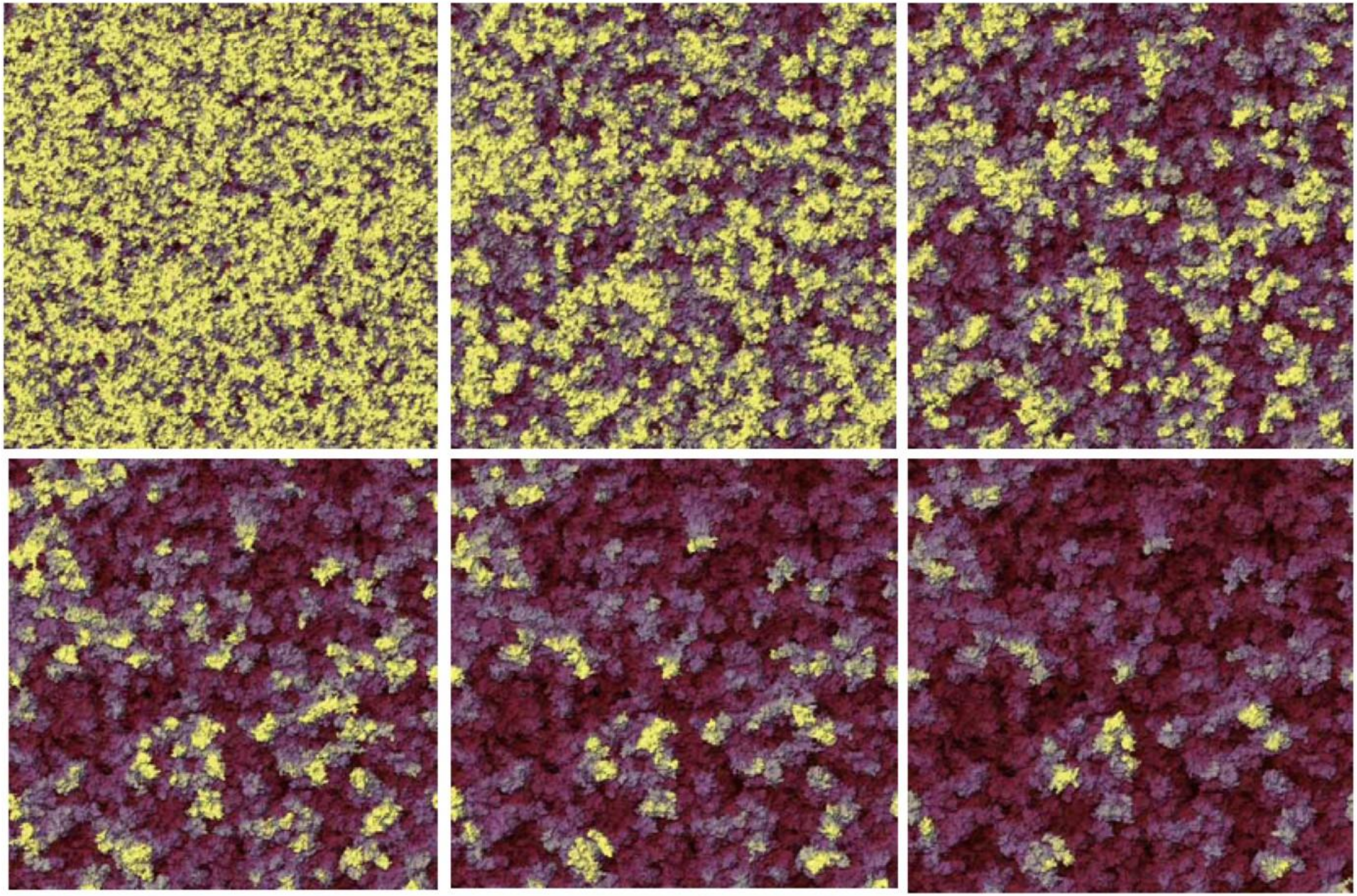


  - By functional-analytic techniques (log Sobolev), dynamics on such a cube mixes in time $O(\log\log n)$ even in L^2-distance.
- Take non-adjacent cubes $Q_1, \dots, Q_N$ ($N \asymp (n/\log^3 n)^d$) and ***imagine*** that the projection on those would predict mixing for the entire system:

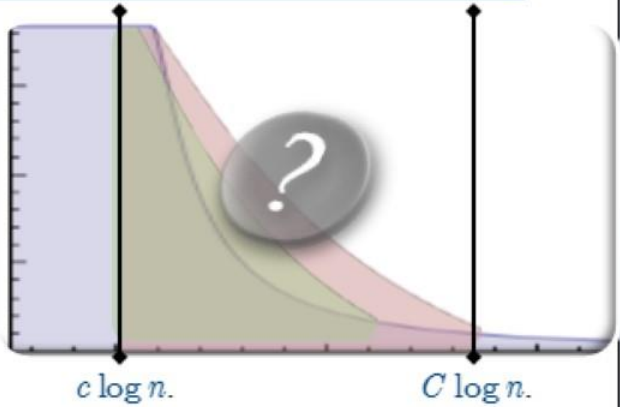


  - Distance between cubes turn them $\approx$ independent.
  - Product of i.i.d.'s with fast L^2-mixing $\rightsquigarrow$ cutoff
  - Expect cutoff at $\frac{1}{2\text{gap}} \log N = \frac{1}{2\text{gap}} \log n + O(\log\log n)$ with window $O(\log\log n)$.

Key tool: breaking dependencies…



Cutoff in expanders



- Recall: Entire convergence of RW on a given family of expanders occurs within $[c \log\ n,\ C \log\ n]$
- What is the typical behavior of regular expanders?

- <u>THEOREM:</u> ([L., Sly '10]:

confirming conjectures of Durrett and Peres

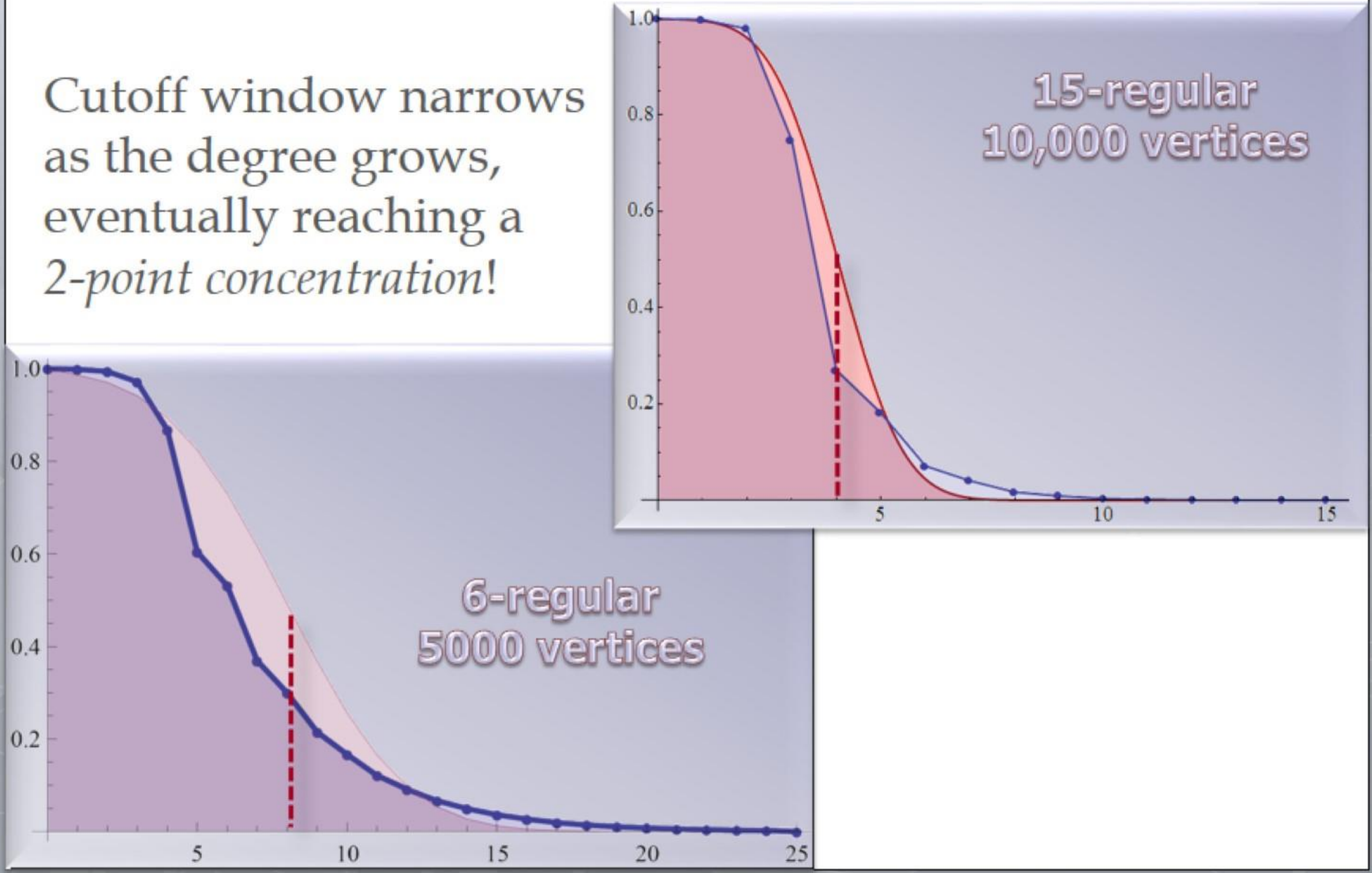
Let G be a random d-regular graph for $d \geq 3$ fixed. W.h.p. the **SRW** on G has cutoff at $\frac{d}{d-2}\log_{d-1} n$ with window $O\left(\sqrt{\log n}\right)$.

- E.g., for almost every random cubic graph:

$$t_{\text{mix}}\left(\tfrac{1}{1000}\right) - t_{\text{mix}}\left(e^{-1}\right) \approx 13.486\sqrt{\log_2 n}$$

Simulations of RWs

Cutoff window narrows as the degree grows, eventually reaching a *2-point concentration*!

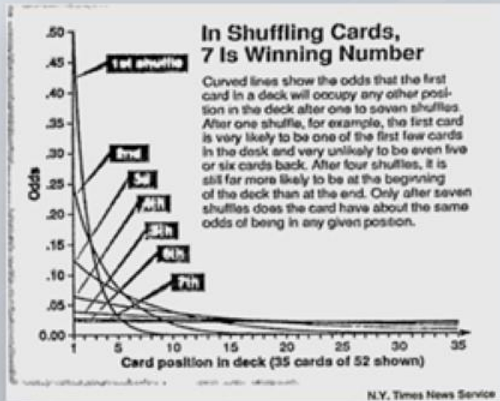


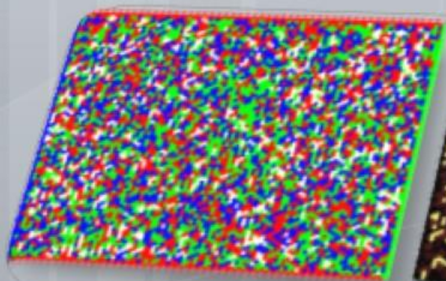
Open problems

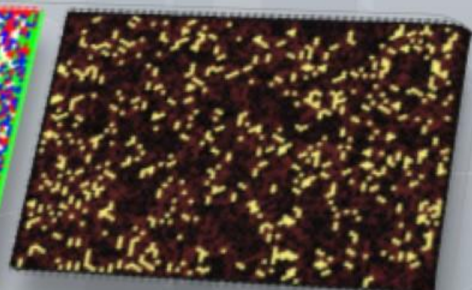
- *Does* RW *exhibit cutoff on* ***every*** *family of transitive 3-regular expanders?*

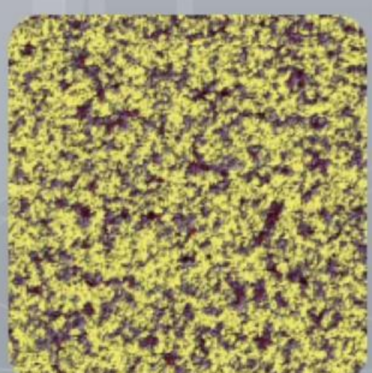
- *Does* RW *exhibit cutoff on* ***any*** *family of transitive 3-regular expanders?*

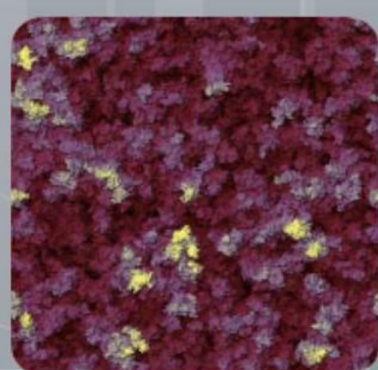
- *Establish cutoff for Metropolis for colorings on a d-regular expander.*











Hebrew University

Math Colloquium

December 2012

Thank you